SOME PRACTICAL ASPECTS OF OVERRELAXATION IN DIAGNOSTIC AND FORECAST CALCULATIONS

JAMES H. S. BRADLEY
The Pennsylvania State University,† University Park, Pa.

ABSTRACT

If the conditions for the existence of a unique optimum overrelaxation factor are satisfied in a jury problem, then component suppression is a useful aid. In jury-marching forecast models, there is no unique optimum overrelaxation factor, because the error vector at acceptance is not dominated by the eigenvector corresponding to the spectral radius. The criterion for fewest iterations is not then the minimization of the spectral radius. It is advantageous to decrease the overrelaxation factor in a predetermined manner during the first two or three time steps or after component suppression. Theoretical results are derived to explain the original purely empirical observations that the coefficient which minimizes the spectral radius requires about 50 percent more computing time than is necessary.

1. INTRODUCTION

Relaxation (iterative or successive approximation) methods are necessary in many finite difference calculations, either for economy or because alternative methods exceed the storage capacity of existing computers. All the procedures discussed in this paper involve a diagnostic problem (also called a jury problem); the forecast procedures involve both jury and marching problems.

A diagnostic problem means finding the solution vector $\mathbf{X}$ of an equation

$$\mathbf{AX}=\mathbf{b} \tag{1}$$

in which the coefficient matrix $\mathbf{A}$ (or order N) and vector $\mathbf{b}$ are known; if the matrix $\mathbf{A}$ is sparse (most of the coefficients zero) or if its elements can be computed by some formula, then it is usually not stored explicitly. An example of a jury problem is the solution of the omega equation, discussed by Stuart and O'Neill [9], and O'Brien [8]; another example occurs in the barotropic model, in which the solution vector $\mathbf{X}$ is the tendency field $\partial\psi/\partial t$ of the stream function ψ and the vector $\mathbf{b}$ represents the rate of change of relative vorticity $\nabla^2(\partial\psi/\partial t)$. Further examples are given in textbooks on numerical weather prediction or numerical analysis, and are included in modern courses in dynamic and synoptic meteorology: Fox [6], Thompson [10], Young [13].

Some initial guess vector $\mathbf{X}^{(0)}$ is chosen; an all-zero vector is commonly used unless some prior information is available. The successive approximations $\mathbf{X}^{(k)}$ may be generated by the Jacobi process

$$x_i^{(k+1)}=-\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ij}x_j^{(k)}+b_i \tag{2}$$

or by the Gauss-Seidel process (which also halves storage)

$$x_i^{(k+1)}=y_i^{(k+1)}=-\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)}-\sum_{j=i+1}^{N} a_{ij}x_i^{(k)}+b_i, \tag{3}$$

or by the Accelerated Liebmann or Overrelaxation process (which requires the same storage as Gauss-Seidel)

$$x_i^{(k+1)}=\omega y_i^{(k+1)}+(1-\omega)x_i^{(k)} \tag{4}$$

where the vector $\mathbf{Y}$ is given by the Gauss-Seidel equation (3), and where ω is the overrelaxation factor. In any of these methods, nearly all the elements a_{ij} are zero. The balance of this paper is concerned solely with the theory and practice of overrelaxation. Practical details of iteration can be found in almost any textbook of numerical analysis, e.g. Fox [6 or 7], Fadeeva [4], Fadeev and Fadeeva [3], Forsythe and Wasow [5], Young [13].

2. EXISTENCE OF AN OPTIMUM

The existence of an optimum overrelaxation factor (called "optimum omega" by many authors) for fewest iterations depends on a property of the error vector $\mathbf{E}^{(k)}$ and residual vector $\mathbf{R}^{(k)}$ which will now be developed.

The error vector

$$\mathbf{E}^{(k)}=\mathbf{X}-\mathbf{X}^{(k)} \tag{5}$$

where $\mathbf{X}$ is the true solution of equation (1), which can be estimated but not determined exactly.

The residual vector

$$\mathbf{R}^{(k)}=\mathbf{X}^{(k+1)}-\mathbf{X}^{(k)} \tag{6}$$

and is known explicitly at every iteration. The usual method is to accept the kth successive approximation $\mathbf{X}^{(k)}$ when some norm $\delta^{(k)}$ of $\mathbf{R}^{(k-1)}$ (for instance, its

†Publication No. 68-2 of the Mineral Industries Experiment Station, The Pennsylvania State University. Present affiliation: McGill University, Montreal, Quebec.

absolutely largest element, or the sum of the squares of its elements, or the sum of the absolute values of its elements) becomes smaller than some given convergence criterion ϵ

$$\delta^{(k)}<\epsilon. \tag{7}$$

Carré [2] and many textbooks show that the iterative procedure (4) is equivalent to

$$\mathbf{E}^{(k)}=\mathbf{M}\mathbf{E}^{(k-1)}=\mathbf{M}^k\mathbf{E}^{(0)} \tag{8}$$

where the relaxation matrix $\mathbf{M}$ is given by

$$\mathbf{M}=-(\omega\mathbf{L}+\mathbf{I})^{-1}[\omega\mathbf{R}+(\omega-1)\mathbf{I}] \tag{9}$$

and where $\mathbf{L}$ is the lower triangular matrix of $\mathbf{A}$, $\mathbf{R}$ is the upper triangular matrix of $\mathbf{A}$ (both $\mathbf{L}$ and $\mathbf{R}$ have zeros on the main diagonal), and $\mathbf{I}$ is the identity matrix (unity on the main diagonal, and zero everywhere else). Equation (9) assumes that $\mathbf{A}$ has ones on the main diagonal, but this causes no loss of generality.

The relaxation matrix $\mathbf{M}$ possesses N scalar eigenvalues λ and N eigenvectors $\mathbf{U}$ which obey

$$\mathbf{M}\mathbf{U}_i=\lambda_i\mathbf{U}_i \tag{10}$$

or equivalently

$$\sum_{l=1}^{N} m_{jl}u_{il}=\lambda_i u_{ij} \tag{11}$$

where u_{ij} is the jth element of $\mathbf{U}_i$. The explicit evaluation of eigenvalues and eigenvectors is discussed in many textbooks (Fox [6 or 7], Fadeev and Fadeeva [3], Wilkinson [12], Todd [11]) but is normally prohibitive for the matrices of high order which occur in diagnostic problems; this paper requires only certain theoretical properties. We shall ignore pathological cases which do not arise in practice.

Any vector $\mathbf{Y}^{(0)}$ may be expressed in terms of the eigenvectors $\mathbf{U}$ and a coefficient vector $\mathbf{C}$ as

$$\mathbf{Y}^{(0)}=c_1\mathbf{U}_1+c_2\mathbf{U}_2+\ \ldots\ +c_N\mathbf{U}_N. \tag{12}$$

Therefore

$$\mathbf{Y}^{(k)}=\mathbf{M}^k\mathbf{Y}^{(0)}=\mathbf{M}^{k-1}(\mathbf{M}\mathbf{Y}^{(0)})$$

$$=\mathbf{M}^{k-1}(\lambda_1c_1\mathbf{U}_1+\lambda_2c_2\mathbf{U}_2+\ \ldots\ +\lambda_Nc_N\mathbf{U}_N) \tag{13}$$

$$=\lambda_1^kc_1\mathbf{U}_1+\lambda_2^kc_2\mathbf{U}_2+\ \ldots\ +\lambda_N^kc_N\mathbf{U}_N \tag{14}$$

from equations (8) and (10).

If λ_1 is the absolutely largest of the eigenvalues λ, and if we ignore the pathological case

$$c_1=0 \tag{15}$$

then for all sufficiently large k

$$|\lambda_1^kc_1|>>|\lambda_i^kc_i| \tag{16}$$

for $i=2,\ 3,\ \ldots,\ N$.

Because an eigenvector may be multiplied by any constant without invalidating equation (10), equation (14) may be used as the basis of the power method (Fox [7], Fadeev and Fadeeva [3], Wilkinson [12]) of estimating the eigenvector $\mathbf{U}_1$ corresponding to the absolutely largest eigenvalue (spectral radius) λ_1 of the matrix $\mathbf{M}$ (or any other matrix), and the spectral radius itself. The simple extensions of (16) necessary when (15) holds are discussed by Fadeev and Fadeeva [3], and Wilkinson [12], but are inconsequential to this paper.

Further, in the particular case when

$$\mathbf{Y}^{(0)}=\mathbf{E}^{(0)}=\mathbf{X}-\mathbf{X}^{(0)} \tag{17}$$

equation (14) shows that the necessary and sufficient condition for the equivalence of minimizing the spectral radius of $\mathbf{M}$ and of minimizing the number of iterations required for acceptance of $\mathbf{X}^{(k)}$ according to equation (7) is that at acceptance the error vector $\mathbf{E}^{(k)}$ shall be dominated by the coefficient $\lambda_1^kc_1$ of $\mathbf{U}_1$.

This assumption is implicit in all the publications cited; it is not normally valid in practical forecast models. When it is invalid, the least number of iterations before acceptance does not necessarily depend on the minimization of any one eigenvalue (or any assortment of eigenvalues); there is not necessarily any unique optimum value of omega for fewest iterations.

An extreme case of purely illustrative interest is that in which the acceptance criterion ϵ is so large that almost any initial guess vector $\mathbf{X}^{(0)}$ is accepted immediately. Practical illustrations are given in section 6.

Nothing in this existence proof has made any assumptions about the matrix $\mathbf{A}$ (except that pathological cases are ignored); in particular, we do not need the assumptions of symmetry, positive definiteness, and "property A" which Carré [2], O'Brien [8], and others must make in order to give an explicit value for the optimum omega in terms of the eigenvalues of $\mathbf{A}$.

These concepts are readily extended to line, block, and alternating-direction methods.

3. AITKEN EXTRAPOLATION

Under the necessary and sufficient condition (16), equations (5) and (8) reduce to

$$\mathbf{X}-\mathbf{X}^{(k)}=\mathbf{E}^{(k)} \tag{5}$$

$$=\mathbf{M}\mathbf{E}^{(k-1)} \tag{8}$$

$$\simeq\lambda_1\mathbf{E}^{(k-1)}=\lambda_1(\mathbf{X}-\mathbf{X}^{(k-1)}). \tag{18}$$

Both $\mathbf{X}^{(k)}$ and $\mathbf{X}^{(k-1)}$ are known; the solution of (18) for $\mathbf{X}$ is called Aitken extrapolation, which is a special form of component suppression (Fadeev and Fadeeva [3]):

$$\mathbf{X}\simeq\tilde{\mathbf{X}}=(\mathbf{X}^{(k)}-\lambda_1\mathbf{X}^{(k-1)})/(1-\lambda_1). \tag{19}$$

Process (19) requires an estimate of the spectra radius λ_1, which is available from the classical theory (Carré [2]).

Part of Carré's work consists effectively in an application of (18) and (19) at three successive iterations to obtain an estimate of the spectral radius λ_1 by the power method; other difficult cases are treated extensively by Fadeev and Fadeeva [3]. It is, however, interesting to consider a generalized Aitken extrapolation based on a more precise form of (18) and on the recognition that condition (16) for the classical theory may be invalid: we shall also begin to distinguish between real and complex eigenvalues.

More precisely, if λ_1 in equation (19) is replaced by an arbitrary constant λ (real if the solution vector $\mathbf{X}$ is real), then the residual error vector

$$\tilde{\mathbf{E}}=\mathbf{X}-\tilde{\mathbf{X}}=(\mathbf{X}-\lambda\mathbf{X}-\mathbf{X}^{(k)}+\lambda\mathbf{X}^{(k-1)})/(1-\lambda) \tag{20}$$

$$=(\mathbf{E}^{(k)}-\lambda\mathbf{E}^{(k-1)})/(1-\lambda) \tag{21}$$

$$=\sum_{i=1}^{N}(\lambda_i-\lambda)\lambda_i^{k-1}c_i\mathbf{U}_i/(1-\lambda) \tag{22}$$

$$=\sum_{i=1}^{N}q_i\lambda_i^{k-1}c_i\mathbf{U}_i \tag{23}$$

where the vector of amplification factors $\mathbf{Q}$ is given by

$$q_i=(\lambda_i-\lambda)/(1-\lambda). \tag{24}$$

Carré [2] proves that the absolutely largest eigenvalue is real and positive for the class of matrix considered in the classical theory if

$$1<\omega<\omega_{opt}. \tag{25}$$

If λ is an estimate of λ_1, the amplification factors will be near zero not only for λ_1 but also for real subdominant eigenvalues close to λ_1. Generalized Aitken extrapolation is a valuable technique when several eigenvectors contribute substantially to the error vector, especially when the spectral radius of the relaxation matrix $\mathbf{M}$ is very close to unity for all values of omega.

It is possible to obtain a reasonable idea of the eigenvalues (provided they are real) corresponding to those eigenvectors which dominate the error vector by trying various values of the extrapolation constant λ. After each extrapolation, the residual error vector is dominated by eigenvectors corresponding to eigenvalues not close to λ. It is inconsequential at what step Aitken extrapolation is used; Carré [2] and Fox [6] recommend its use at the end of a relaxation solely in order to obtain λ as an estimate of λ_1 by the power method.

4. COMPLEX EIGENVALUES

Practical relaxation matrices have some or all of their eigenvalues and eigenvectors complex (Carré [2]), even though the matrix $\mathbf{A}$, the solution vector $\mathbf{X}$, the vector $\mathbf{b}$ and all intermediate calculations involve only real quantities.

If $\mathbf{A}$ is symmetric, then the complex eigenvalues of $\mathbf{M}$ occur in complex conjugate pairs. When $\mathbf{A}$ is not symmetric (for instance, in the National Meteorological Center operational barotropic model with mountains and friction), the complex eigenvalues and eigenvectors are not necessarily complex conjugate; nevertheless all the calculations may involve only real quantities.

In the complex case, all references to absolute values must be replaced by references to moduli.

According to Fadeev and Fadeeva [3], methods of component suppression analogous to generalized Aitken extrapolation are poorly developed in the complex domain; the powerful methods available in the real domain require increased variable storage.

5. FORECAST MODELS

Many finite difference forecast models known to the author involve a jury problem of the form of equation (1); nearly all these models involve a jury problem at each time step. In a wide class of geostrophic, quasi-geostrophic and stream function models, there is a jury problem in which the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are known functions of the state of the system (for instance, in the simplest barotropic models $\mathbf{A}$ is a fixed matrix and $\mathbf{b}$ is the time derivative of the relative vorticity), and the solution vector $\mathbf{X}$ is the time derivative of the geopotential or stream field. Some models require repeated solutions of the vertical velocity (omega) equation discussed by Stuart and O'Neill [9] and O'Brien [8].

The significant difference from ordinary jury problems is in the first guess $\mathbf{X}^{(0)}$ at the solution vector. In forecast models, it is common to use

$$\mathbf{X}^{(0)}_{t=0}=0 \tag{26}$$

at time step $t=0$, but at each subsequent time step to use as the first guess the solution vector accepted at the previous time step.

$$\mathbf{X}^{(0)}_{t=T}=\mathbf{X}^{(accepted)}_{t=T-1}. \tag{27}$$

In practical cases, the condition

$$|\lambda_1^k c_1|>>|\lambda_i^k c_i| \quad , \quad c=2, 3, \ldots, N \tag{16}$$

is satisfied only at the first one or two time steps. At all subsequent time steps, the convergence criterion

$$\delta^{(k)}<\epsilon \tag{7}$$

is satisfied with fewest iterations by minimizing some statistical melange of eigenvalues, eigenvectors, and coefficients, all of which depend in an unknown manner on the overrelaxation factor ω, time, and initial conditions. Empiricism and rules of thumb are the best solutions to such a problem.

In meteorological terms, the interpretation of these phenomena is simple. As noted by Thompson [10], the largest eigenvalues in forecast models are associated with the planetary waves; at time $t=0$ and possibly $t=1$, it is the planetary waves whose tendencies dominate the error vector. At all subsequent times, the tendencies of

TABLE 1.—*Number N of interations required and maximum residual R for the simple barotropic model of Young [13]. The residual is for height tendency $\partial z/\partial t$ in cm./sec. $\times 10^{-3}$; the convergence criterion given by equation (7) was a maximum modulus norm less than 4×10^{-3} cm./sec., which causes differences of less than 1 ft., in a 36-hr. forecast.*

Overrelaxation Factor	Time Step 1		2		3		4		Time Step 16		17		18		19		Time Step 32		33		34		35	
	N	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*	*N*	*R*
1.65	18	3.44	12	2.69	12	2.86	11	3.59	16	3.73	17	3.60	19	2.92	20	3.05	27	2.22	22	3.31	26	2.66	23	3.82
1.58	15	3.92	10	3.10	10	3.53	9	3.96	14	3.00	15	3.24	16	3.48	17	3.32	22	3.47	18	3.31	20	3.57	20	3.24
1.56	15	3.75	10	2.69	10	2.86	9	3.57	13	3.58	14	3.54	15	3.56	16	3.59	21	3.16	17	3.33	19	3.36	19	3.39
1.54	15	3.64	10	2.35	9	3.61	9	3.23	13	3.19	14	3.00	14	3.70	15	3.67	20	3.28	16	3.25	18	3.27	18	3.68
1.52	15	3.64	9	3.39	9	3.35	9	2.91	12	3.89	13	3.52	14	3.05	14	3.39	19	3.44	15	3.35	17	3.33	18	2.96
1.50	15	3.71	9	3.08	9	2.94	8	3.97	12	3.49	13	3.06	13	3.39	13	2.96	18	3.55	14	3.61	16	3.46	17	3.23
1.45	15	3.95	9	2.49	8	4.00	8	2.92	11	3.85	12	3.13	12	3.61	13	3.63	17	2.79	14	3.49	15	2.66	16	2.59
1.42	16	3.60	8	3.53	8	3.21	8	2.43	11	3.34	11	3.85	12	3.78	13	3.79	16	2.85	14	3.22	14	2.84	15	2.97
1.40	16	3.72	8	3.19	8	2.90	7	3.27	11	3.38	11	3.89	12	3.87	13	3.90	15	3.65	14	2.98	13	3.81	14	3.81
1.38	16	3.85	8	2.87	8	2.60	7	3.27	11	3.45	11	3.96	12	3.97	14	3.54	15	2.94	13	3.99	13	3.07	14	3.16
1.36	16	3.97	8	2.56	8	2.32	7	2.90	11	3.51	12	3.48	12	3.97	14	3.62	14	3.65	13	3.57	13	2.40	14	2.55

the slowly moving planetary waves change little, so that equation (27) is a good guess for them; the error vector for $t>1$ is dominated by eigenvectors corresponding to the fast-moving short waves, which are associated with relatively small eigenvalues.

6. VERIFICATION

In this case, the verification preceded and largely inspired the theory. Table 1 shows the number of iterations required at each time step and the maximum residual at acceptance for the 14- by 23-point barotropic model described by Young [13], for which he estimated the optimum omega (for minimum spectral radius) as 1.65.

At the first time step, the optimum omega is near 1.53; at steps 2–4 it is less than 1.36. The experiments were not carried to sufficiently low omega because it was not realized that the optimum omega might depend on the time step. The optimum value fluctuates slightly during the forecast; it is not known how this effect is partitioned between changes in the waves present in the system and accidental fluctuations in the residuals. Comparable figures can be obtained only on identical models; the exact numbers vary with the sequence of iteration and other coding changes inconsequential to the forecast, because of differences in the coefficient vector **C** in equation (13). Experiments with an independently coded version of the same model showed that the optimum omega was usually near 1.35 but sometimes fell as low as 1.20. The computing time is increased about 50 percent by the use of the theoretically optimum value.

Separate experiments were made to verify that the accepted tendency field was essentially independent of the value of the convergence criterion in the working range.

A few experiments were performed on the National Meteorological Center operational barotropic model; although the results were similar and independent of the convergence criterion, they are difficult to analyze theoretically, because the fact that the model sets negative absolute vorticities to zero proved to be a major effect in the cases of unusually bad forecasts which were available as initial data (Bradley, Hayden, and Wiin-Nielsen [1]) and because the mountain and friction terms cause the matrix **A** in equation (1) to vary slowly. The operational model was modified, however, to make the overrelaxation factor a predetermined function of the time step, 1.70 initially and 1.30 at all subsequent steps.

Experiments with Aitken extrapolation by equation (20) were performed on Young's barotropic model for the values of the extrapolation factor λ from 2.0 to -2.0 for overrelaxation factors from 1.8 to 1.05, and for convergence criteria of 0.004 to 0.001 cm./sec. In all cases, a very slight and occasional advantage was obtained for λ in the range 0.1 to -0.1, so small that the additional computations were uneconomic. The extrapolation was performed at every fifth iteration.

This result may be interpreted in terms of Carré's [2] theory of the distribution of eigenvalues; the real eigenvalues are positive and greater than or equal to $(\omega-1)$, and the complex eigenvalues all are of modulus $(\omega-1)$. The results are construed to mean that even at the first time step and even for the tightest convergence criterion (0.001 cm./sec.) the error vector is not dominated by the eigenvector corresponding to the spectral radius according to equation (16): equation (24) shows that the modulus of the amplification factor q cannot necessarily be made less than unity for complex eigenvalues. This interpretation is consistent with the finding that the optimum omega at the first time step is lower than the value of 1.65 predicted by Young [13].

The author intends to include tests of more powerful component suppression methods in a future paper.

ACKNOWLEDGMENTS

I wish to thank Dr. C. Young for theoretical discussions; Mr. L. Norton for coding and running the results in table 1; Miss Mary Ann Ruzecki for the NMC operational barotropic model; and Dr. R. C. F. Bartels for special privileges at the University of Michigan Computing Center.

REFERENCES

1. J. H. S. Bradley, C. M. Hayden, and A. C. Wiin-Nielsen, "An Attempt to Use Satellite Photography in Numerical Weather Prediction," *Report* 07424-1-T, Dept. of Meteorology and Oceanography, University of Michigan, Ann Arbor, June 1966, 72 pp.
2. B. A. Carré, "The Determination of the Optimum Accelerating Factor for Successive Over-Relaxation," *Computer Journal*, Vol. 4, No. 1, British Computer Society, London, Apr. 1961, pp. 73–78.
3. D. K. Fadeev and V. N. Fadeeva, *Computational Methods of Linear Algebra*, Freeman, San Francisco, 1963, 621 pp.
4. V. N. Fadeeva, *Computational Methods of Linear Algebra*, Dover Publications, New York, 1959, 252 pp.
5. G. E. Forsythe and W. R. Wasow, *Finite-Difference Methods for Partial Differential Equations*, John Wiley and Sons, New York, 1960, 444 pp.
6. L. Fox, *Numerical Solution of Ordinary and Partial Differential Equations*, Pergamon Press, Oxford, 1962, 509 pp.
7. L. Fox, *An Introduction to Numerical Linear Algebra*, Oxford University Press, London, 1964, 328 pp.
8. J. J. O'Brien, "Comments on 'The Over-Relaxation Factor in the Numerical Solution of the Omega Equation,'" *Monthly Weather Review*, Vol. 96, No. 2, Feb. 1968, pp. 99–104.
9. D. W. Stuart and T. H. R. O'Neill, "The Over-Relaxation Factor in the Numerical Solution of the Omega Equation," *Monthly Weather Review*, Vol. 95, No. 5, May 1967, pp. 303–307.
10. P. D. Thompson, *Numerical Weather Analysis and Prediction*, Macmillan, New York, 1961, 170 pp.
11. J. Todd, *Survey of Numerical Analysis*, McGraw-Hill Book Co., Inc., New York, 1962, 589 pp.
12. J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, London, 1965, 662 pp.
13. C. Young, *Computer Applications in the Atmospheric Sciences*, Dept. of Meteorology and Oceanography, University of Michigan, Ann Arbor, 1965, 256 pp.

[*Received June 6, 1967; revised and expanded April 19, 1968; revised July 10, 1968*]